

TheAnalyticsTeam

Sprocket Central Pty Ltd

Data analytics approach

[Data Analytics] - [Engagement Manager], [Senior Consultant], [Narendranath Singh]

Agenda

1. Introduction
2. Data Exploration
3. Model Development
4. Interpretation

Introduction

Based on the dataset provided, detail approach for following slides. Our team focus:

1. Understanding Data Distribution
2. Feature Engineering
3. Data Transformation
4. Model building
5. Interpretation

Data Exploration

Understanding Data:

Dataset Customer Demographic:

1. Total rows are 4000 and 13 columns
2. Missing values: Last Name 125 rows with missing last name, 87 rows with missing DOB, 506 missing job title, 656 missing industry category, 87 rows with missing tenures.
3. Inconsistency in values in gender columns, there are values with M,F, Male, Female and U

Dataset Customer Address:

1. The dataset has no missing values.
2. The state defined is not consistent, state is defined as NSW and also New South Wales.

Dataset Transaction:

1. There are 20,000 records of transaction for year 2017.
2. Missing values: Column online_order is missing 360 records with no values.
3. Missing values: Columns brand, product_line, product_class, product_size, standard_cost are missing values for 197 records. If records is missing one column the same row has missing value in other above columns.

Data Exploration

Feature Engineering Dataset:

Merge Datasets: Customer Demographic, Customer Address and Transactions:

1. Dropping row element with extreme DOB value: 1843-12-21, as it is skewed data.
2. Rename Gender values and some values were F/M. Renaming U as 'Unknown'
3. Replacing missing values with 'Unknown' in Job Title and Job Industry Category.
4. Deceased Indicator has most values as N, dropping the column.
5. Created additional column Age using DOB.
6. State columns has ambiguous values, Victoria and New South Wales, converting them to VIC and NSW respectively.

Analysis on Transaction Dataset:

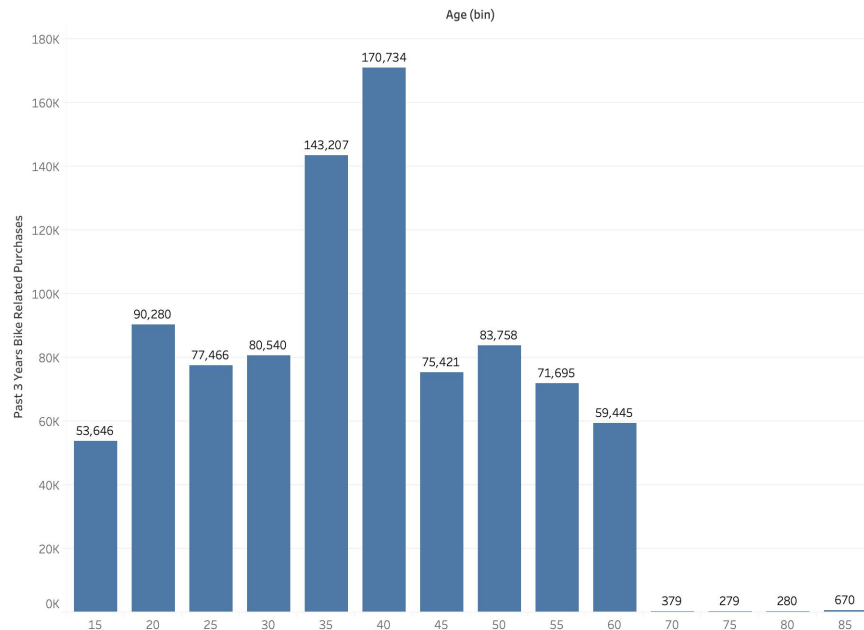
7. Similar values are missing in columns: Brand, Product Line, Product Class and Product Size, dropping these rows.
8. Created additional column Profit using List Price and Standard Cost
9. Created additional column Age using DOB.

Data Exploration

Customer Segmentation based on Age:

Distribution of people based on age groups, we can observe from the data that total number of rental purchases of bike for past 3 years are between age group 30 and 45.

Distribution of 3 year purchase between age groups.



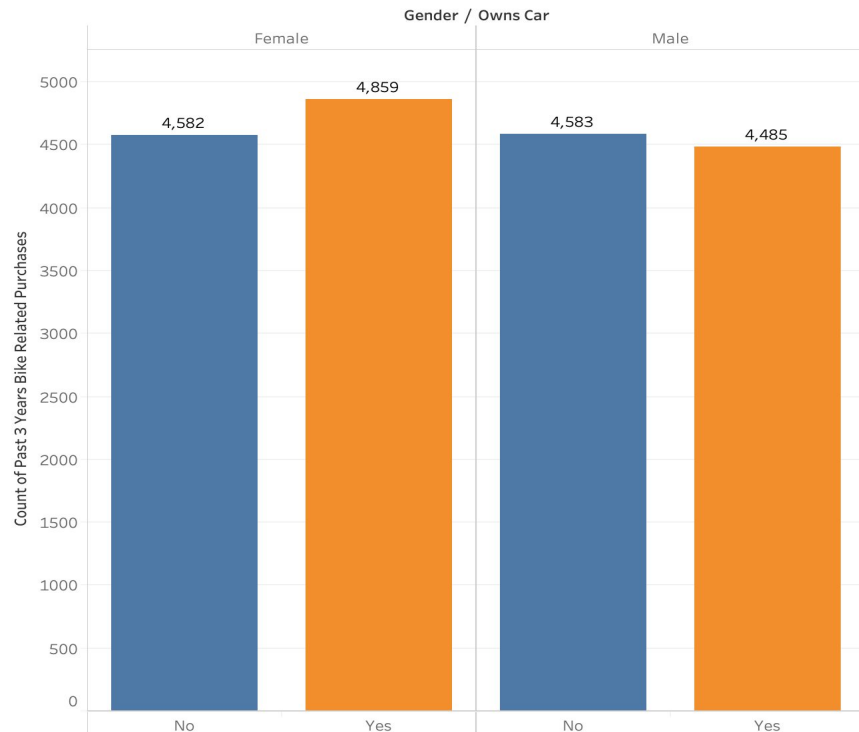
Data Exploration

Customer Segmentation based on Gender:

Distribution of people based on gender group, we can observe from the data of past 3 years of purchases:

1. Females are more likely to rent the bike than men, females have rented total 9,441 times compared to men 9,068 times.
2. Also, females with cars are more likely to rent a bike than men, females with cars have rented total of 4,859 to men 4,485.
3. There are no difference in rental of bikes if they don't own a car.

Purchases made based on own a car

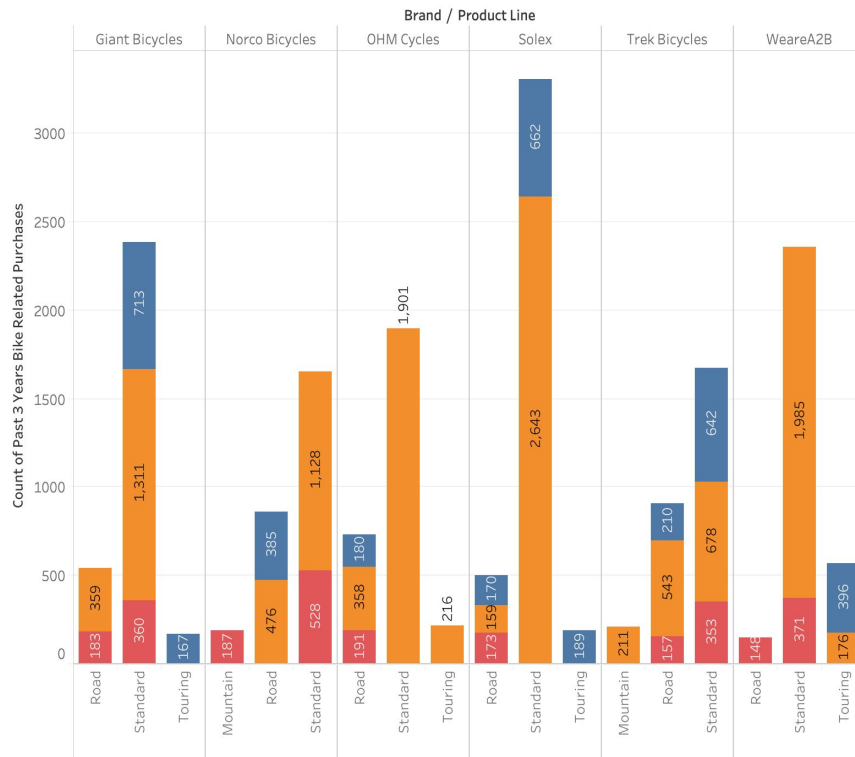


Data Exploration

Brand Segmentation based on 3 year rental:

1. The most popular brand among the people is Solex Standard Product line with total of 3,305 times rented.
2. Overall for all the brands standard is most popular product line.

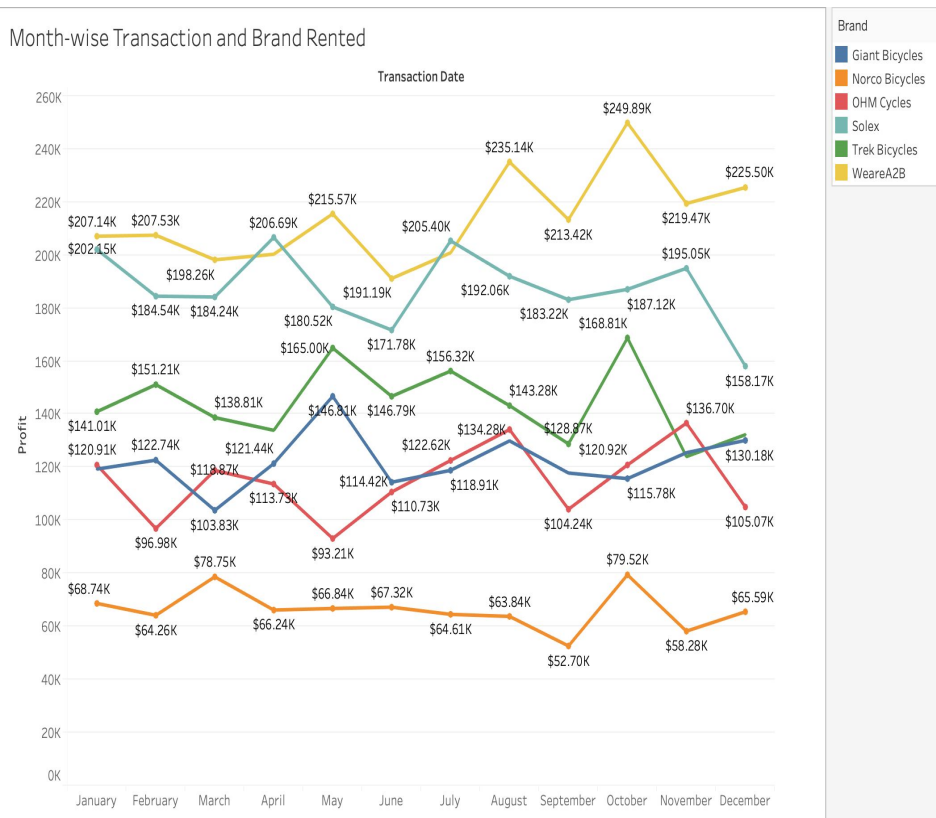
Past 3 year purchases for brand and product line.



Data Exploration

Monthwise distribution Profits:

1. The most profitable brand is WeareA2B.
2. The most profitable month is October for for year 2017 with sales profit of \$922.03K.



Data Exploration

Hotspots Area of most Profitable area:

1. New South Wales and Victoria Area is more profitable rental area for bikes.

Hotspot area for most profitable area

